

Diferencias en Diferencias Apiladas y Proyecciones Locales

Jorge Pérez Pérez¹

¹Banco de México

Conferencia de Stata Latinoamérica
Sep 13, 2023

El contenido del documento y sus conclusiones son responsabilidad exclusiva de los autores y no reflejan necesariamente la opinión del Banco de México.

Motivación

- ▶ El método de diferencias en diferencias es una herramienta común para evaluar los efectos de políticas económicas.
- ▶ Tradicionalmente, estos modelos se estiman mediante regresiones de efectos fijos bidireccionales (“Two-way fixed effects”).
- ▶ Una literatura reciente señala que estos estimadores pueden tener problemas cuando los efectos de las políticas son heterogéneos entre unidades (Goodman-Bacon, 2021; Callaway and Sant’Anna, 2021; De Chaisemartin and d’Haultfoeuille, 2020).
- ▶ Varios estimadores disponibles para recuperar el efecto promedio de políticas en escenarios de heterogeneidad en los efectos.

En esta presentación

- ▶ Dos enfoques alternativos para estimar efectos de políticas en escenarios en donde estos efectos son heterogéneos:
 - ▶ Diferencias en diferencias apiladas (Cengiz et al., 2019).
 - ▶ Proyecciones locales (Dube et al., 2023).
- ▶ Implementación en Stata 18.
- ▶ Comparación de estos enfoques con los comandos para estimación de efectos tratamiento heterogéneos en Stata 18.

CONFERENCIAS STATA LATAM 2023

Aplicaciones estadísticas al servicio
de distintas áreas del conocimiento

Contexto



STATA

Diferencias en diferencias



Fuente: Manual de Stata 18

$$Y_{i,t} = \alpha + Tratado_i + Post_t + Tratado_i \times Post_t + \varepsilon_{i,t}$$

Diferencias en diferencias

$$Y_{i,t} = \alpha + \text{Tratado}_i + \text{Post}_t + \theta \text{Tratado}_i \times \text{Post}_t + \varepsilon_{i,t}$$

En un escenario con múltiples unidades, posiblemente con tratamientos no simultáneos, era usual la estimación con efectos fijos bidireccionales (TWFE)

$$Y_{i,t} = \alpha_i + \delta_t + \beta T_{i,t} + \varepsilon_{i,t}$$

β : Efecto tratamiento promedio entre los tratados (ATT)

Diferencias en diferencias y estudio de evento

Para estimar efectos de tratamiento que varían en cada momento del tiempo, la especificación usual es la de estudio de evento en panel

$$Y_{i,t} = \alpha_i + \delta_t + \sum_k \beta_h \mathbf{1}(\text{tiempo de evento}=h) + \varepsilon_{i,t}$$

donde el tiempo de evento es el número de periodos que han pasado desde que la unidad fue tratada.

Diferencias en diferencias con efectos tratamiento heterogéneos

$$Y_{i,t} = \alpha_i + \delta_t + \beta T_{i,t} + \varepsilon_{i,t}$$

$$Y_{i,t} = \alpha_i + \delta_t + \sum_k \beta_k 1(\text{tiempo de evento}=k) + \varepsilon_{i,t}$$

- Literatura reciente ha mostrado que en contextos con efectos tratamiento heterogéneos, el coeficiente β estimado de la regresión de efectos fijos bidireccionales no corresponde con el ATT (Goodman-Bacon, 2021; Callaway and Sant'Anna, 2021; De Chaisemartin and d'Haultfoeuille, 2020).

Diferencias en diferencias con efectos tratamiento heterogéneos

- ▶ Si hay diferentes grupos de observaciones $g = 1, \dots, G$, cada uno con un efecto tratamiento diferente que varía en el tiempo, el efecto tratamiento para el grupo g a horizonte h es:

$$\beta_{g,h} = E(y_{i,h}(1) - y_{i,h}(0)|g)$$

- ▶ El efecto tratamiento promedio en los tratados es

$$ATT_h = E[\beta_{g,h}]$$

Diferencias en diferencias con efectos tratamiento heterogéneos

$$Y_{i,t} = \alpha_i + \delta_t + \sum_h \beta_h 1(\text{tiempo de evento}=h) + \varepsilon_{i,t}$$

Problema: β_h no es un estimador insesgado de ATT_h

- ▶ Solo surge si los efectos tratamiento son diferentes entre grupos
- ▶ O si el tratamiento ocurre en diferentes periodos para diferentes unidades

Múltiples soluciones en la literatura

- ▶ Callaway and Sant'Anna (2021)
- ▶ Wooldridge (2021)
- ▶ Comandos `hdidregress` y `xthdidregress` en Stata 18.
- ▶ **Diferencias en diferencias apiladas** (Cengiz et al., 2019)
- ▶ **Proyecciones Locales** (Dube et al., 2023)

Diferencias en diferencias apiladas

Diferencias en diferencias apiladas

- ▶ Los problemas de la estimación de TWFE solo ocurren en diseños con múltiples unidades tratadas y efectos heterogéneos. No ocurren en comparaciones entre dos unidades, una tratada y una de control.
- ▶ Supongamos que tenemos un grupo de control $g = 1$ y múltiples grupos tratados $g = 2, \dots, G$
- ▶ ¿Podemos convertir el problema de estimación en comparaciones entre dos unidades? Sun and Abraham (2021):
 1. Estimar muchos efectos tratamiento: uno para cada grupo
 2. Cada uno de los efectos se estima con diferencias en diferencias con dos unidades: las unidades de cada grupo tratado y las unidades del grupo de control $g = 1$
 3. Luego promediar estos efectos (ponderando por el número de observaciones en cada grupo) para obtener ATT

Diferencias en diferencias apiladas

Muchas estimaciones en una sola regresión:

- ▶ Crear muchas bases de datos, una base por grupo tratado, incluyendo la unidad tratada y todas las unidades de control
- ▶ Apilar todas las bases de datos
- ▶ Estimar TWFE con efectos fijos de base de datos por unidad y base de datos por tiempo
- ▶ Inferencia debe hacerse usando Ferman and Pinto (2019) (No la cubro hoy).

CONFERENCIAS STATA LATAM 2023

Aplicaciones estadísticas al servicio
de distintas áreas del conocimiento

Proyecciones locales



STATA

Proyecciones locales

- ▶ En la estimación apilada:
 - ▶ Se mantienen la unidad tratada y las unidades de control en cada caso
 - ▶ Se estima un efecto por grupo (cohorte de tratamiento)
 - ▶ Se estiman efectos dinámicos (para cada horizonte de tratamiento)
- ▶ Las bases de datos apiladas podrían generarse para cada grupo y cada horizonte de tratamiento
- ▶ La estimación en cada base de datos correspondería a una estimación 2x2: control vs. un grupo tratado, pre vs. post

Proyecciones locales

Dube et al. (2023) proponen el siguiente enfoque. Estimar la siguiente regresión para cada horizonte h :

$$y_{i,t+h} - y_{i,t-1} = \beta^h \Delta T_{i,t} + \delta_t^h + \epsilon_{i,t}^h$$

Restringiendo las observaciones a aquellas para las cuales:

- ▶ $\Delta T_{i,t} = 1$ (Tratadas nuevas)
- ▶ $T_{i,t+h} = 0$ (Unidades de control)

Proyecciones locales

$$y_{i,t+h} - y_{i,t-1} = \beta^h \Delta T_{i,t} + \delta_t^h + \epsilon_{i,t}^h$$

- ▶ Evita el problema de heterogeneidad al excluir unidades previamente tratadas
- ▶ Usa solo unidades no tratadas como control
- ▶ Fácil de calcular, fácil para hacer inferencia en “bootstrap”

Conclusiones

Conclusiones

- ▶ Múltiples enfoques para estimar efectos de política mediante diferencias en presencia de efectos heterogéneos
- ▶ Enfoques de diferencias en diferencias apiladas y proyecciones locales:
 - ▶ Usan comparaciones de unidades tratadas contra unidades nunca tratadas
 - ▶ Usan herramientas tradicionales de regresión
 - ▶ Fáciles y rápidos de implementar

Referencias I

- Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of econometrics*, 225(2):200–230.
- Cengiz, D., Dube, A., Lindner, A., and Zipperer, B. (2019). The effect of minimum wages on low-wage jobs. *The Quarterly Journal of Economics*, 134(3):1405–1454.
- De Chaisemartin, C. and d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–2996.

Referencias II

- Dube, A., Girardi, D., Jorda, O., and Taylor, A. M. (2023). A local projections approach to difference-in-differences event studies. Technical report, National Bureau of Economic Research.
- Ferman, B. and Pinto, C. (2019). Inference in differences-in-differences with few treated groups and heteroskedasticity. *Review of Economics and Statistics*, 101(3):452–467.
- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.
- Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.

Referencias III

Wooldridge, J. M. (2021). Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators. *Available at SSRN 3906345.*